Making a Successful Strategy game for the Apple App Store

Jiabing Ruan
Exploratory Data Analysis

I. BACKGROUND

Motivation

It is difficult to become a develop a top game in the Apple App Store. There is an abundance of games, and, with the number of phones purchased increasing, a market is created in which demand and supply are both high There are many factors that go into creating a successful app. Two factors that we considered as primary indicators of success were popularity and quality. The indicator variables we chose to represent these factors were average user rating and the total number of user ratings for an app. In this report, we explore the relationship between these two indicator variables and other variables within our data set that we believed could have an impact on the success of an app. We hope that this information could be beneficial to developers of apps and investors of the companies that develop apps.

Industry Overview

Mobile gaming is a billion dollar industry. There are currently 2.7 billion smartphone users worldwide, and 194 billion app downloads in 2018 alone. For this reason, it is necessary for companies to spend large amounts of money and resources on the development, publication and marketing of games. Strategic game development has now become a must in an industry that is becoming increasingly saturated.

Data Set

We explored a variety of datasets in search of data that met the needs of our project. We prioritized finding one with a relatively large sample size, appropriate/informative variables, and recent data. The dataset we decided on can be found on Kaggle. It contains 17,007 games that were scraped/collected on 3rd August, 2019, through the iTunes API and App Store sitemap.

At the time of import, the dataset contained 18 variables, all of which have been defined below:

URL	A link that sends you directly to the game on the app store.
ID	An ID assigned to the game
Name	The official name of the game on the app store
Subtitle	Secondary text under the name
Icon URL	A link to the game's display picture

Average User Rating	Rounded to the nearest .5 and requires at least 5 ratings to make the list
User Rating Count	Number of ratings internationally. null means it is below 5
Price	Price in USD
In-app Purchases	Prices of available in-app purchases
Description	App description
Developer	App developer
Age Rating	Either one of 4+, 9+, 12+ or 17+
Languages	ISO2A language codes
Size	Size of the app in bytes
Primary Genre	The main genre of the game
Genres	Other genres of the game
Original Release Date	When the game was released
Current Version Release Date	When the game was last updated

Exploration paths

We took a few different paths when exploring the data. The first was to manipulate the variables in our data set so that we could look for the relationships within the data we were interested in. For example, we created dummy variables and categorical variables out of some of the numeric data within the data set. We also used different types of plots to explore different relationships within the data. The variables we created during the data cleaning process were helpful in this step.

II. DATA PRE-PROCESSING

Data Cleaning

Before started analyzing our data set, we needed to ensure that the data was in a usable form and clean it otherwise. Some of the games fell into a category we dubbed as “useless games” because they had no user ratings. We also dropped two variables (columns) that we thought did not serve any purpose in our project: 'url' and 'icon_url'.

Upon further exploration of the dataset, we discovered that many of our variables contain NA values. These variables were: subtitle, price, in_app_purchases, and languages. For the “subtitle”, we did not see the need to fill those NA values. For the “price”, we replaced the NA

values with a 0. For the in_app_purchases and languages, we decided to hold off on dealing with the NA values temporarily because of a plan to apply more complex ways of preprocessing to the column, in the next phase of the project.

Data Manipulation

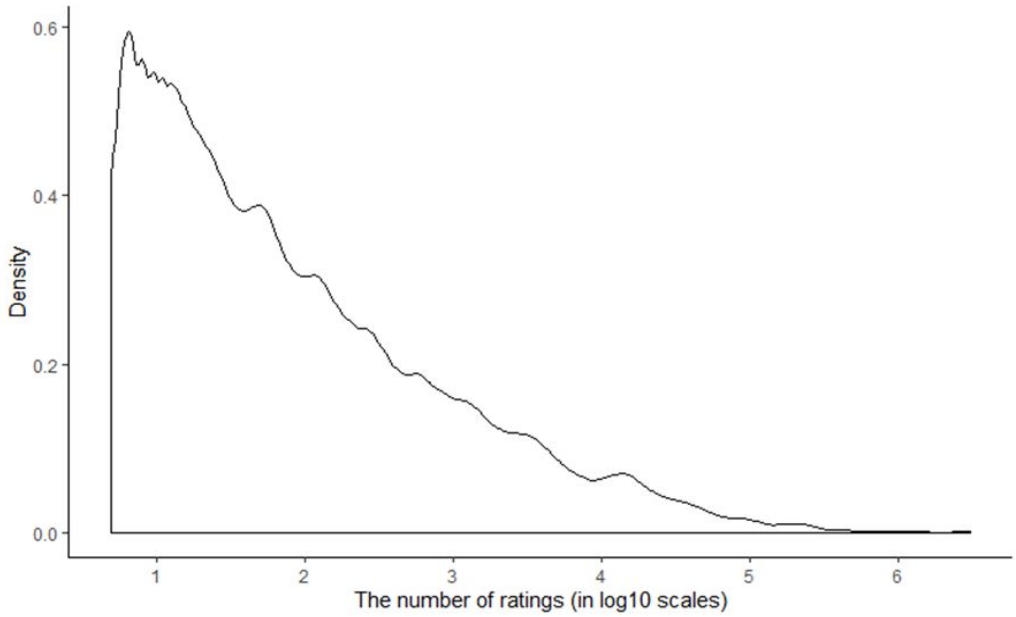
- We first processed the language variable, which contains all the supported languages of a game in the same index (i.e. "EN, JA, RU"). We pivoted the long columns into wider columns, thereby each language becomes an independent dummy variable. Since there are more than 110 languages, we only kept the top 10 most frequently supported languages. (Also because there is a sharp decrease of counts from 10th to 11th)
- We extracted release year from the original variable: original release date, because we needed to explore the data set by year. This method is also applied when we explored the current version release year and extracted only the year values.
- Moreover, we transformed the variable in app purchases into a dummy variable. We think the presence of in-app purchases alone may have larger impact on customers compared to the individual prices on the in-app purchases, because one game may have multiple in app purchases options.
- Since there are multiple subgenres for one game, we extracted the most important subgenre of each game by choosing the one listed first.
- We transformed the size of games from byte into size_MB which is in megabytes. This was done to decrease the magnitude of data.

III. DATA EXPLORATION

Initial Thoughts

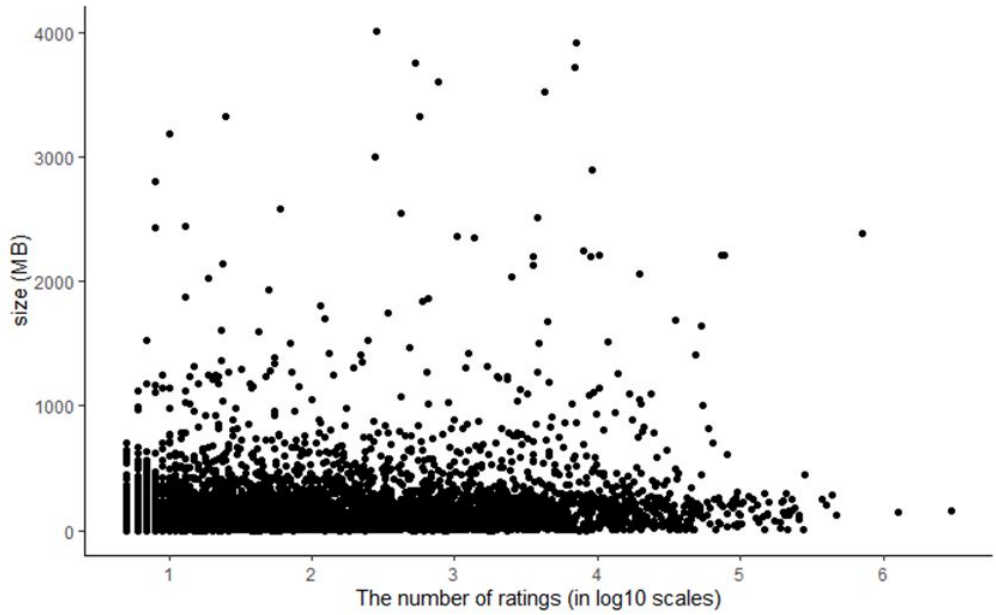
For this data set, our primary indicator of success is the total number of ratings. We are interested in what elements of a game impact its popularity. We also observe what elements impact a game's quality through our other success indicator, average user rating. This will be done in the deeper exploration.

We noticed that there is a large amount of variance within the total number of user rating variable. For this reason we decided to isolate the top 10 percent of the games within this variable, and compare them with the overall market of games available in our data set. For example, we observed how the size of the top 10 percent of games has changed over the years compared to the size of all of the games. We decided that the outliers, in this situation, would be considered successful by our definition. The graphs below explains our decision for analyzing the top 10 percent:



This graph depicts the density of total ratings within our data set. We used the log 10 scale for the number of ratings. This scale is used to account for the skewness towards large values(i.e., cases in which one or a few points are much larger than the bulk of the data). Even though we use this scale, we still see a much higher density for lower number of ratings, which means that most games have a relatively low number of ratings. In other words, the outliers on the graph are actually the much more popular and successful games. We could also see large gaps between the games with a high number of ratings and fewer number of ratings. Popular games are difficult to create. This is why there is not a high proportion of them within this data set.

This next plot shows the relationship between size and the number of ratings. Larger size games have more details, better graphs and rich contents. Therefore, we assume that popular games will have larger size in general. Scale is also set to log 10 since we are using all the data. However, from this scatter plot, there is no significant relationship shown between the size and the number of ratings. We believe the results may be influenced also by year, and propose the idea of using top 10 percent to find this relationship. Then we will make a comparison between the top 10 percent and all games.



Another reason for choosing the top 10 percent is that there is a significant change in the lowest number of ratings when we looked at the top 10, 15, and 20 percent. Values are shown below in the table: we can see that the lowest number of ratings at the bottom for each group changes more than two times from the group of top 15 percent to the group of top 10 percent. In contrast, the lowest number of observations changes less from the group of top 20 percent to the group of top 15 percent. Along with the analysis that we get from the above two graphs, we have chosen to test on the top 10 percent of the games based on the number of ratings.

Top Percentage	Lowest Number of Observations
10%	2512
15%	1088
20%	557

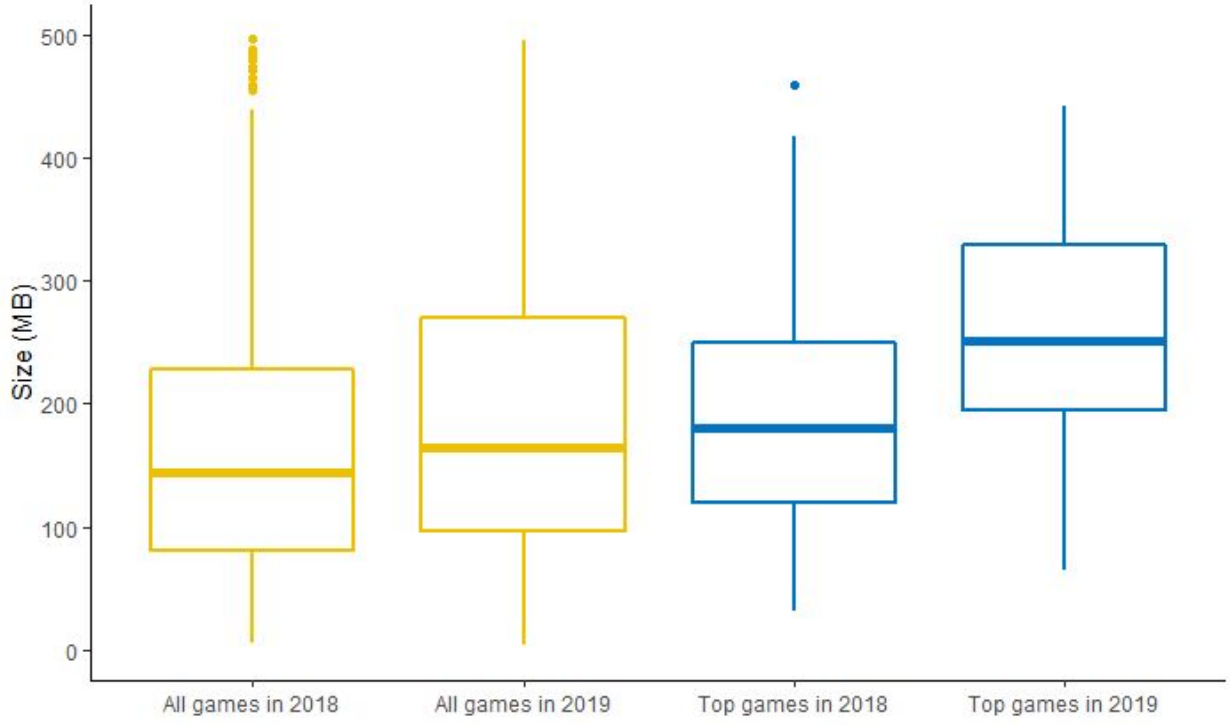
Variable Exploration

- **Size**

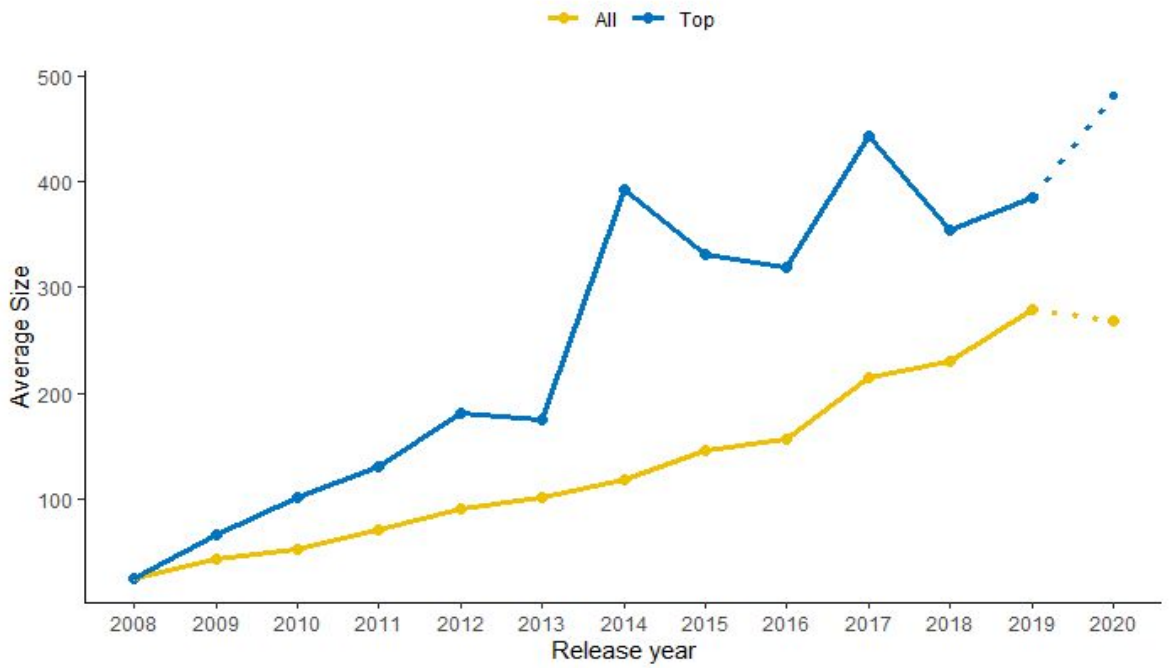
The average sizes of games is largely impacted by outliers in our dataset. The median size is also lower than the value for the mean size in certain cases. We decided to plot the graphs of all games and top games side by side to get a better idea of what the ideal size of a game is.

The graph below informs us that the most popular games tend to have more content, details or better graphics. This explains why they generally have larger sizes.

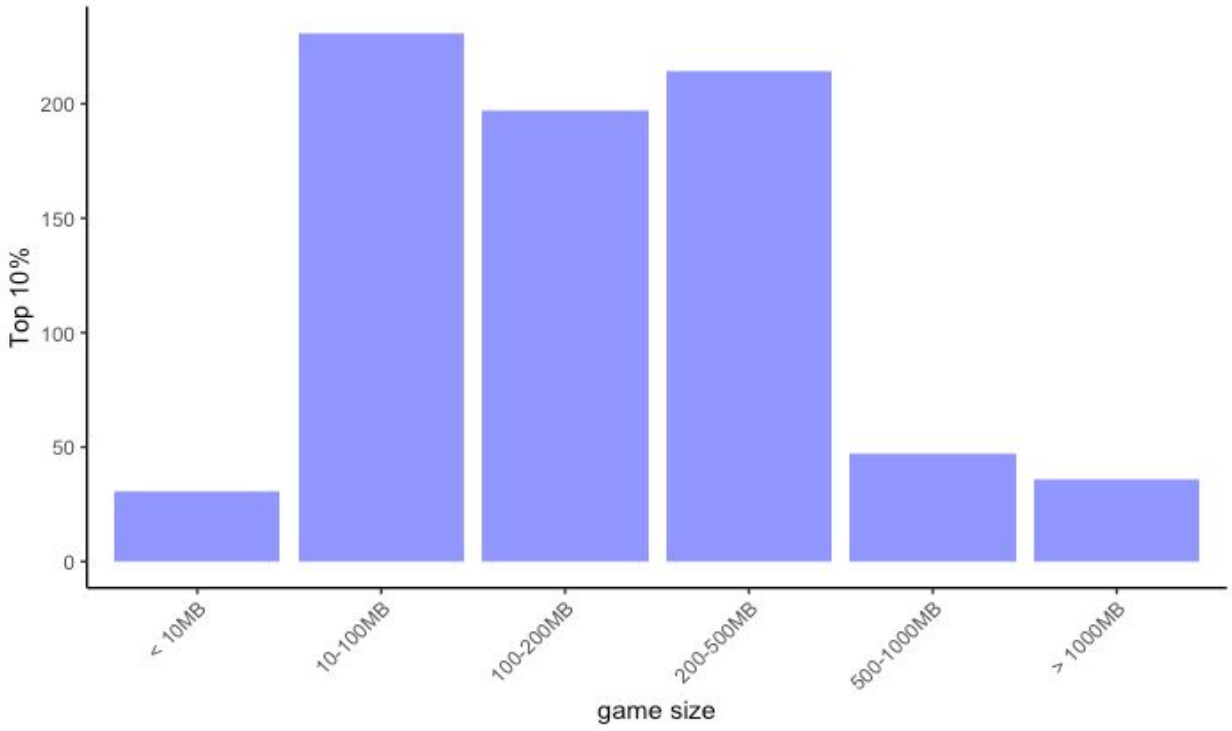
It is plausible to suspect that mobile games that are less than 200 MB could be lacking in certain aspects that make the game enjoyable. According to the trend and the distribution, 250-450MB may be an ideal size for a mobile game.



There is an increasing trend in the average sizes of strategy games on the app store, year after year. We performed a T-test that revealed a significant difference in the mean distribution of the size of these two groups. A Linear regression model we investigated also predicts that the average size of top 10% of games may approach the 450 MB size range by 2020. The T-test also revealed a significant difference between the averages for all and top games.

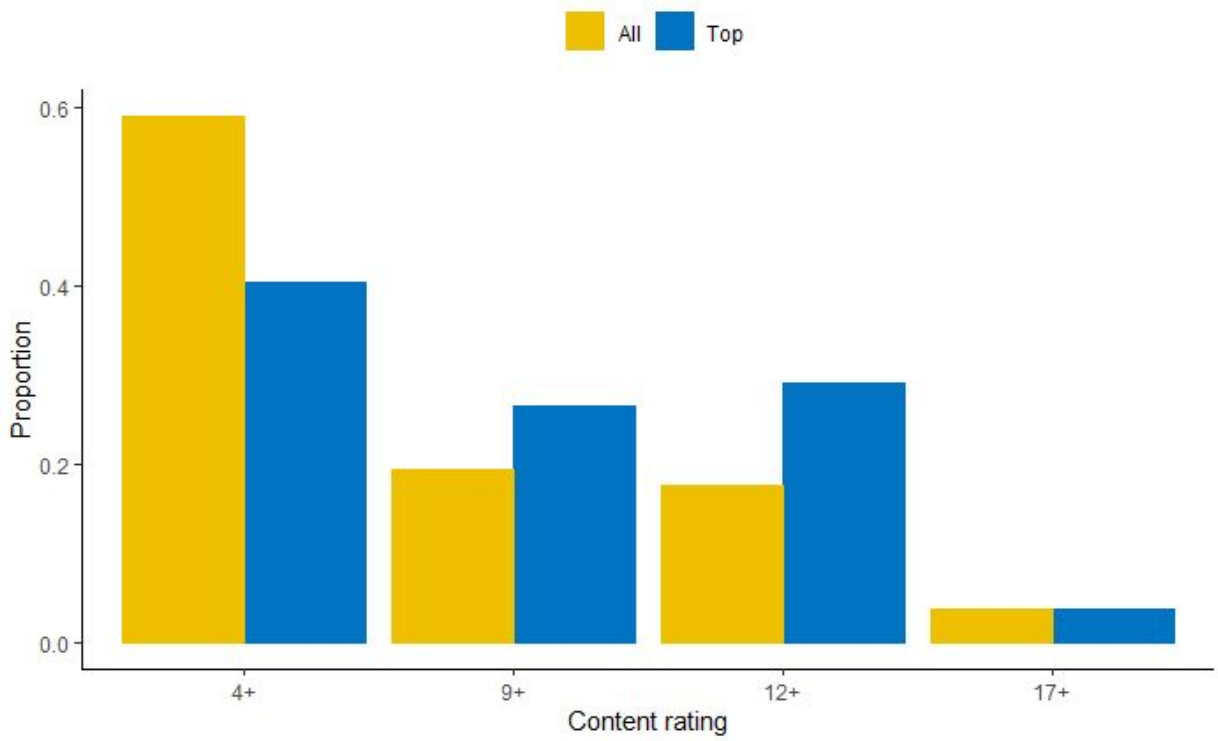


After recognizing that popular games tend to have more content, hence their larger sizes, we decided to explore the sizes within the top 10% of games. The result supports our earlier claim about the sizes of games. We see that the majority of games fall within the 100-500 MB range.

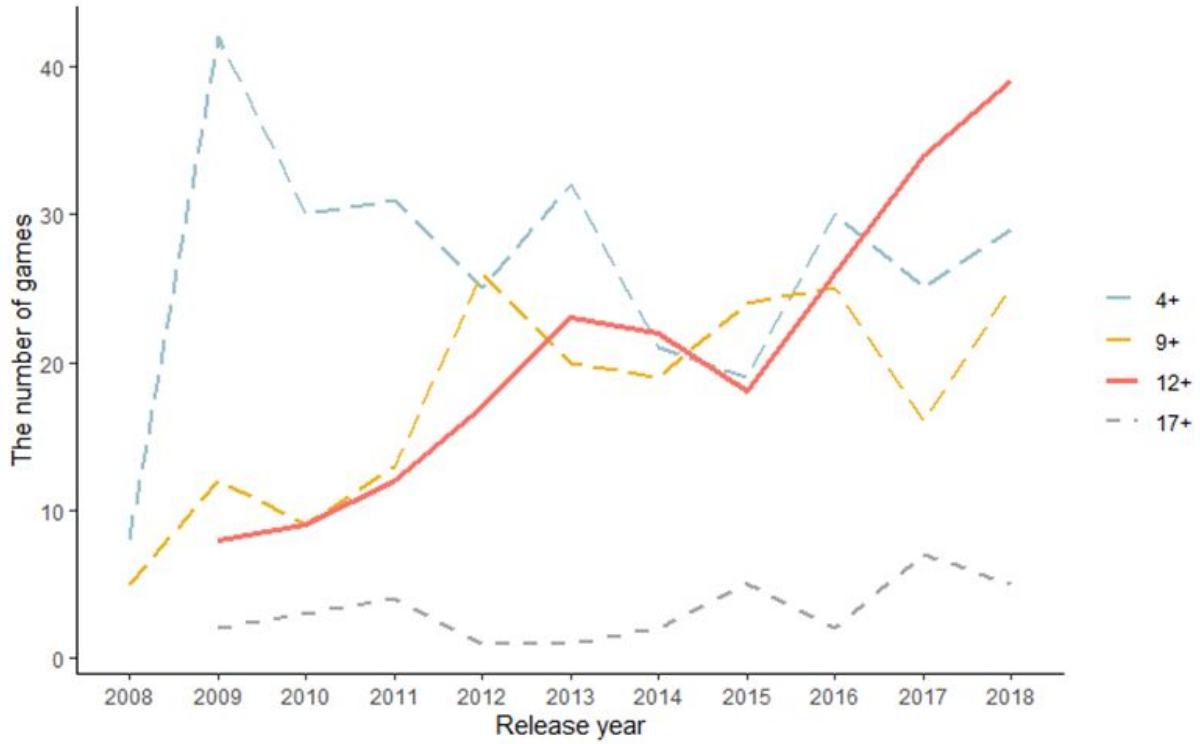


- **Age**

We next explored the proportion of games by content rating (age). We could see that for both groups, most games are rated as “4+” (lowest age for a strategy game). For the all-games group, there is a decreasing trend shown between the four age types. However, for top-10% group, there is an increasing trend shown from “9+” to “12+”. Content rating is based on several categories including violence, sexual or horror elements. We believe those will have an effect on the popularity of games.

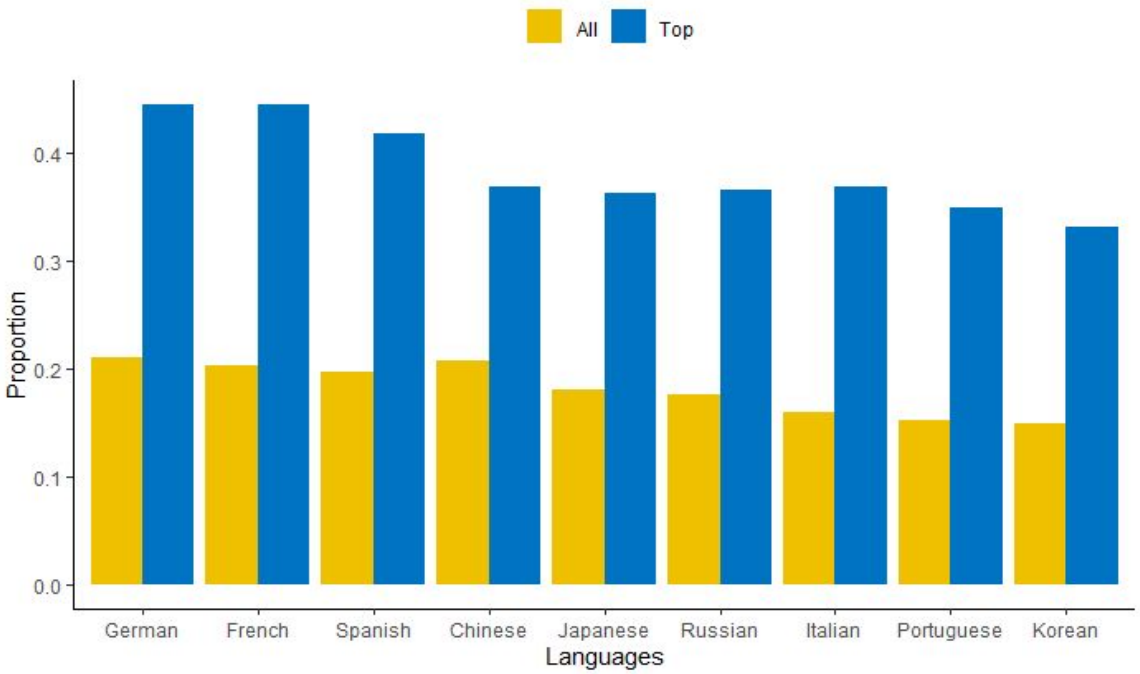


We also observed the yearly trend of the top 10% percent games. There is a significant increase in games that rated as “12+”. The changes for games that rated as “4+” or “9+” fluctuate, but do not show any significant decrease or increase compared to the games that are rated “12+”. We can see that games rated “17+” have relatively low counts. Therefore, developing 17+ games may not be a good idea. Violent or sexual elements may make the customers uncomfortable. We believe that moderate amounts of violent or sexual elements may make the game more playable or interesting. “12+” games would be enjoyed by people from 12 to about 25 years old because they don’t like to play games that make them feel naïve. Moreover, this age range covers the largest group of customers who love to play games.

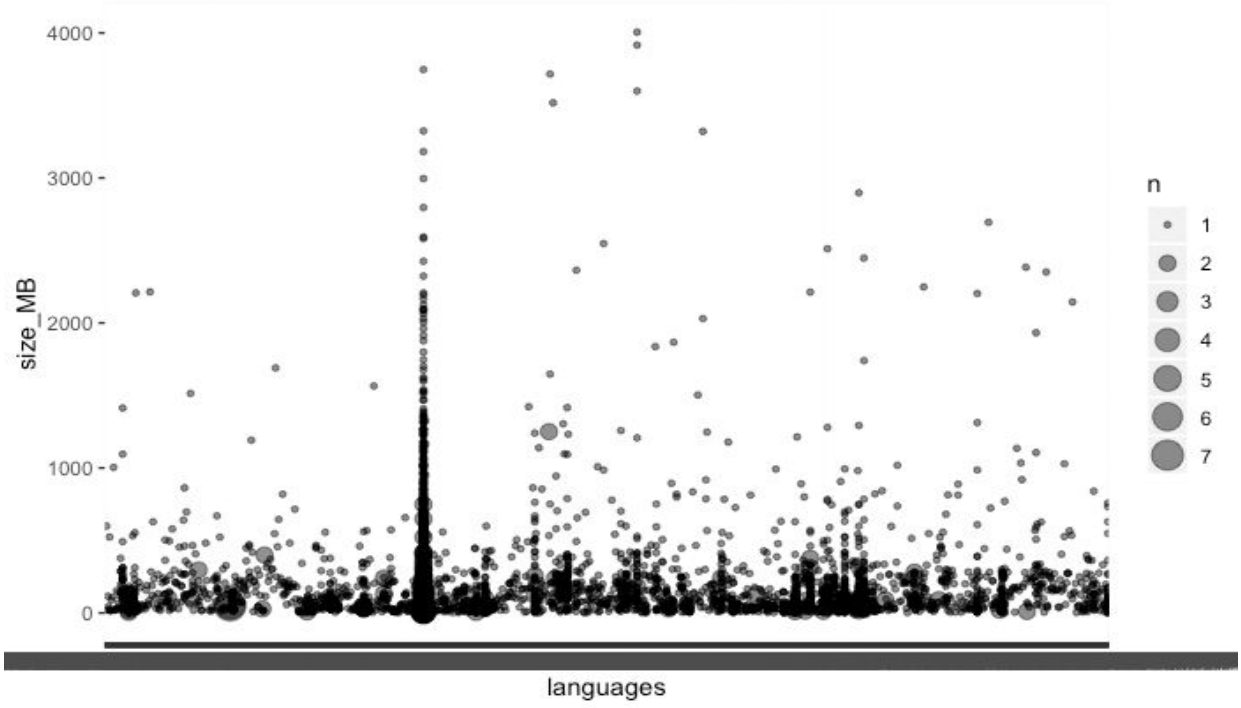


- **Language**

We noticed that almost every game on the App Store supports English. For this reason, we excluded it from the list of languages we explore and analyze for this project. From the graph below, we can see that the top 10% games are more likely to support multiple languages (English included). We also conducted a prop test for each language to confirm that the proportion of top 10% games that support German are significantly different from the proportion of all games that support German. From this analysis, we were able to conclude that a game that supports multiple languages is likely to be popular and successful on the App Store.

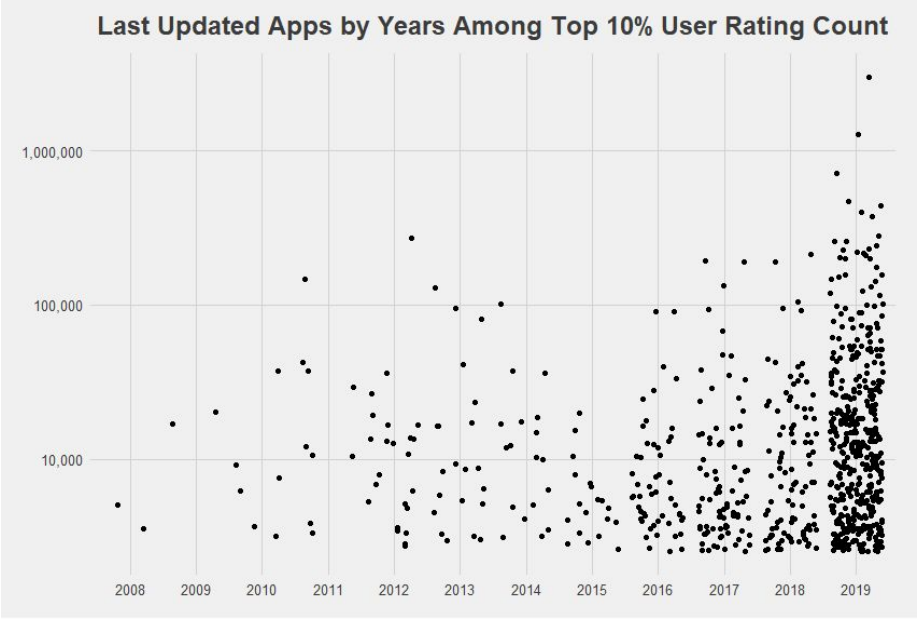


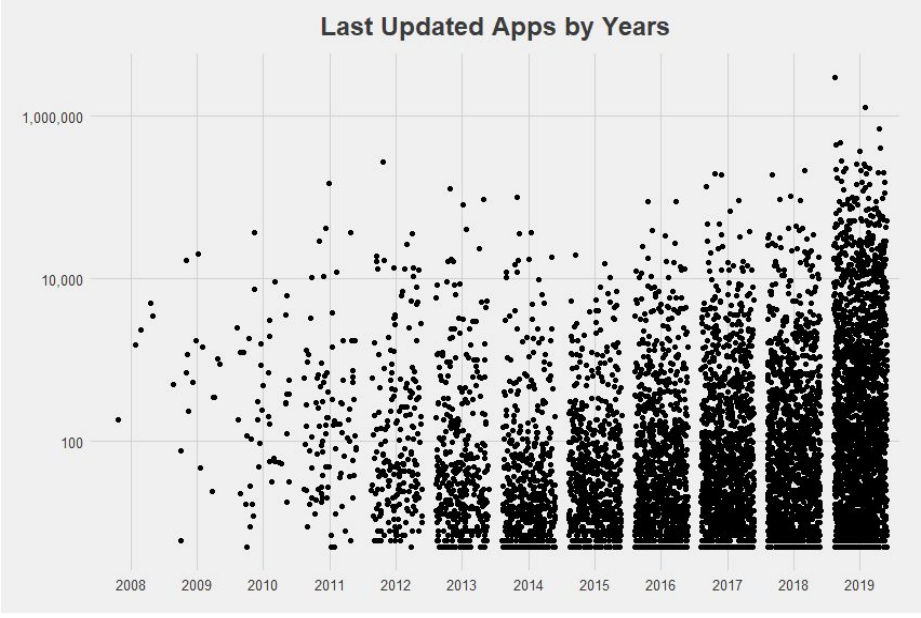
We thought that games that support multiple languages might have larger sizes too, because of our assumption that the top games generally have larger sizes. From this graph, we are able to tell that this is not the case. The games that support the most languages seem to have the smaller sizes, as observed in the graph above. Here, "n" refers to the number of languages a particular game contains.



- **Year Updated**

Here, applied tests to find the relationship between the last time an app was updated (in years) and the number of ratings. We see that for both groups, most games have released their latest version in 2019. However, the difference is not large between those two groups. They both have increasing trends from about 2015 to 2019. The plot shows that more points are aggregating around the year of 2019 in the graph for the top-10% group.

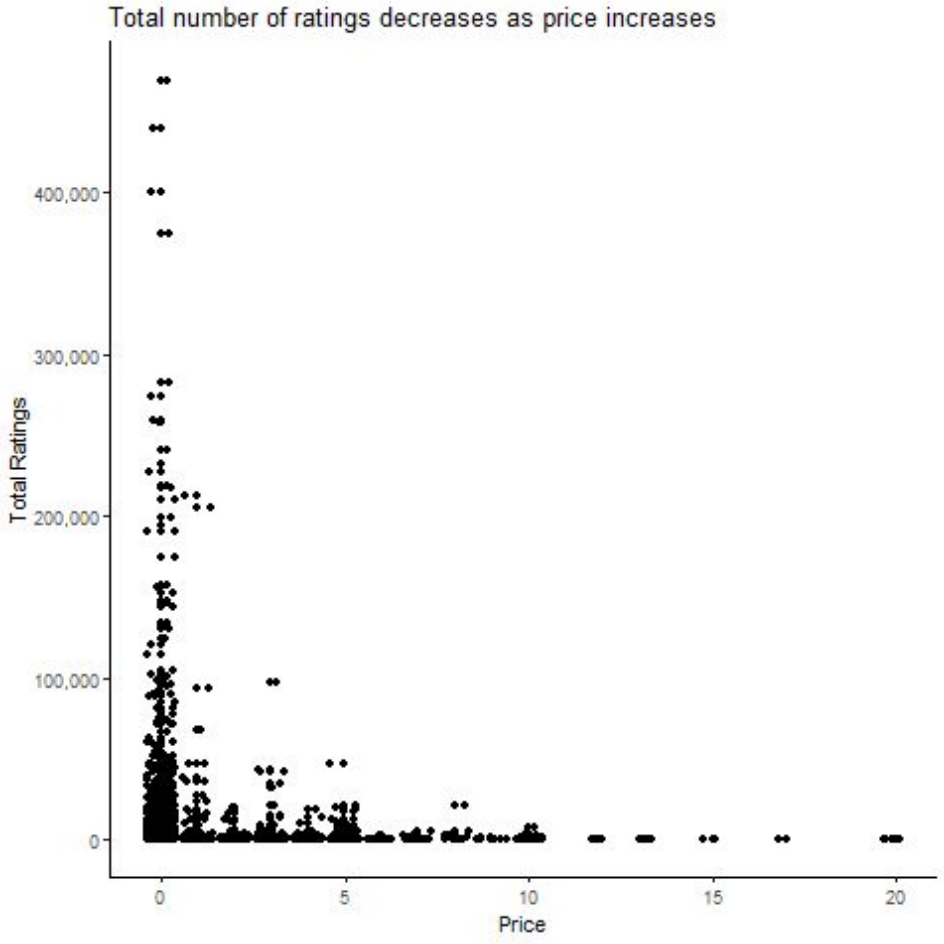




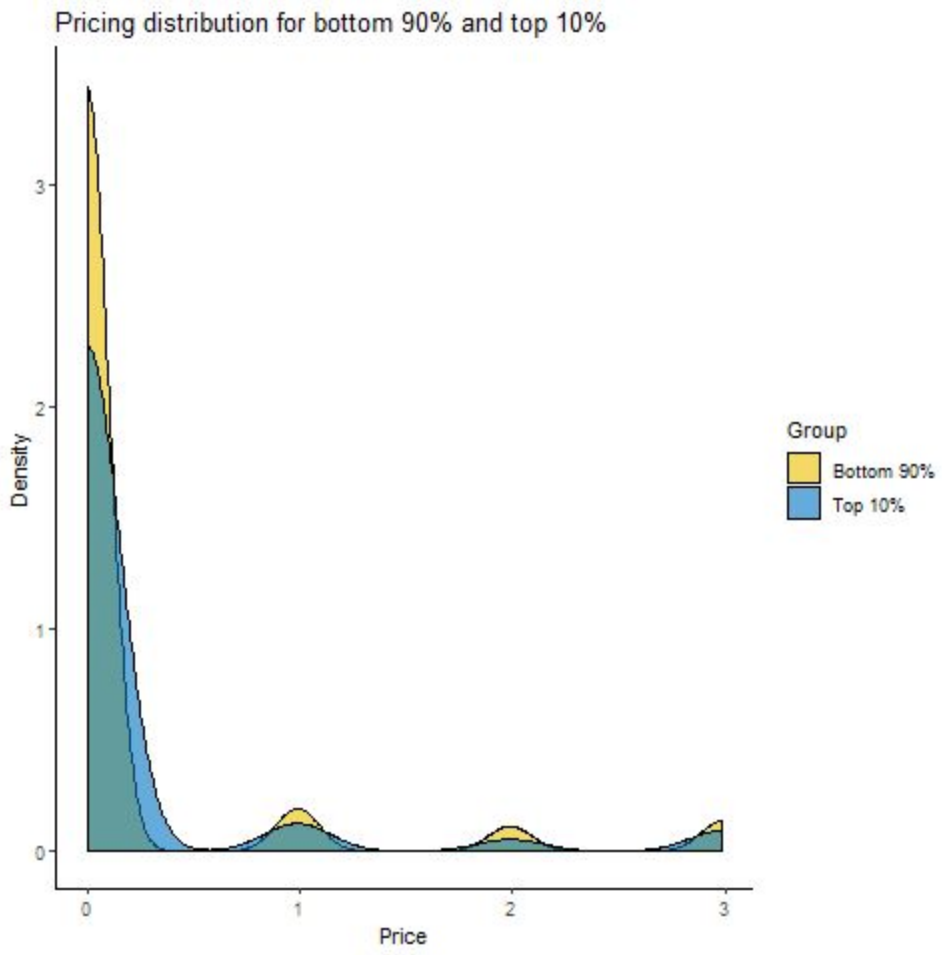
- **Price & In-App Purchases**

Apps can range in price from $0 to $100. Price in this data set refers to the upfront cost of downloading the app. Consumers can also make in-app purchases once the app is downloaded that can enhance the experience of the game. We wanted to look at how price and the presence of in-app purchases changed between the top 10%, the bottom 90%, and the total market for strategy games in the Apple app store.

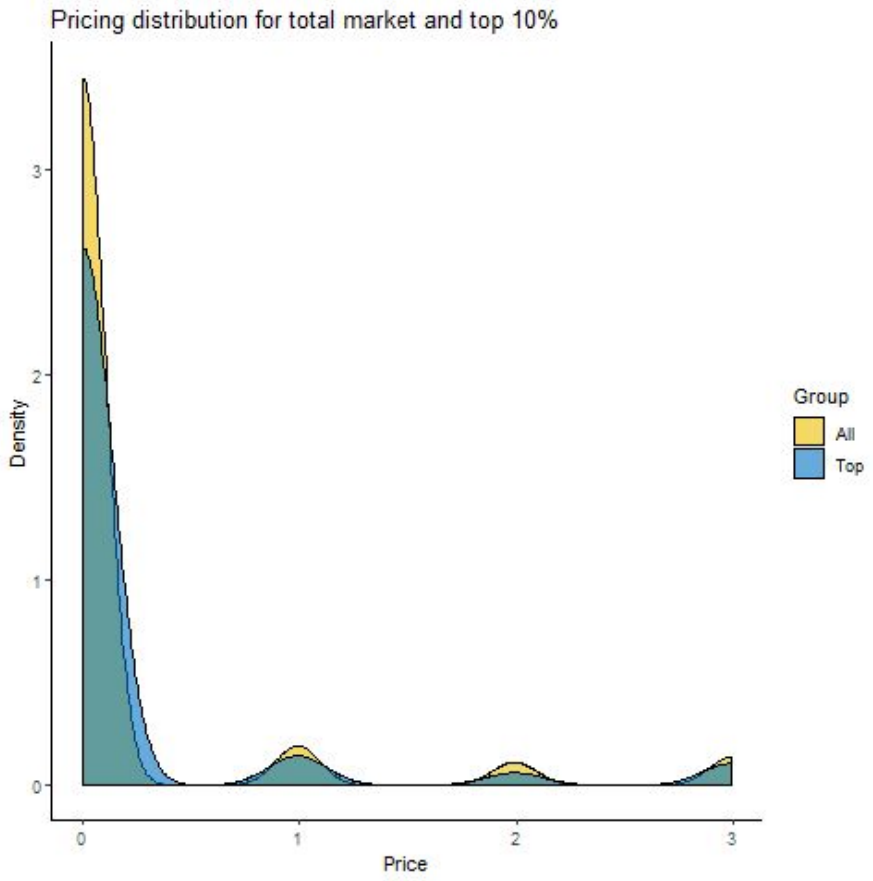
The plot below shows that as price increases, the total number of user ratings decreases. This comes as a result of consumers being more willing to download an app if the price is low or even free.



The pricing distributions we constructed support this claim. We observed comparisons between the top 10% of our data set and the bottom 90% and total market for strategy games. The plot below shows how the pricing distribution of the top 10% of games compared to the bottom 90%.

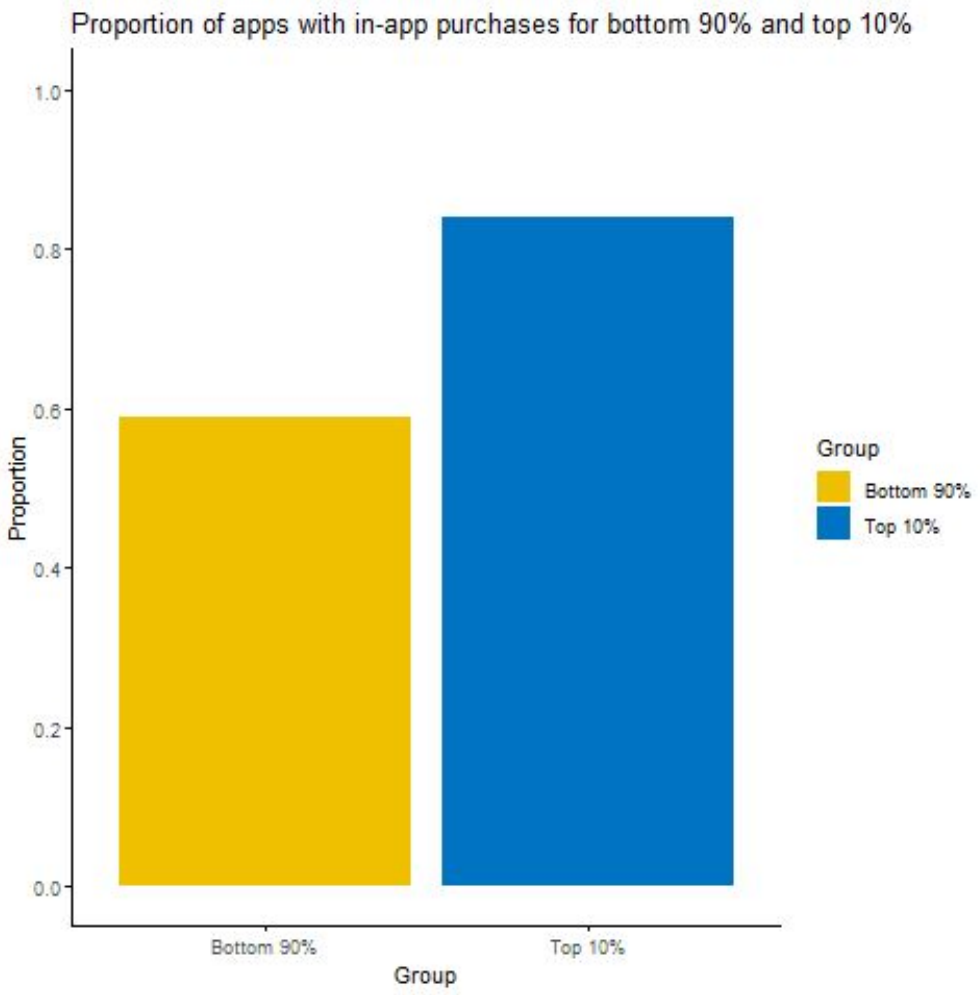


The bottom 95% of the pricing distribution was used here to highlight the density between $0 and $1. The pricing distribution for both groups is relatively the same, with a high density of games priced between $0 and $1. The same relationship can be seen below when the top 10% of the data set is compared to the total market for strategy games. We concluded that, based on comparisons between the two groups, there is no discernible difference in the upfront cost of an app between the top tier games and the rest of the market.

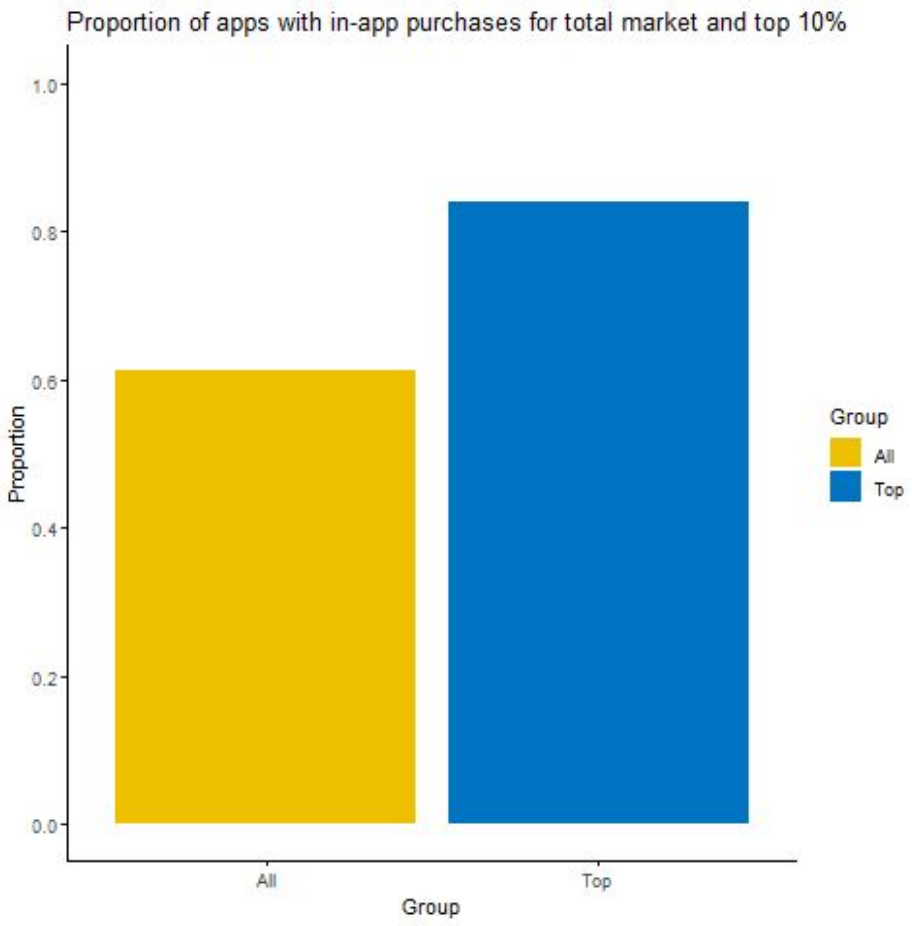


The presence of in-app purchases is also something we wanted to observe. Many apps contain in-app purchases, and we wanted to see if there was any difference between the top games and the rest of the market. Two comparisons were made to view this relationship. The top 10% of games was compared again to the bottom 90% of the market as well as the total market.

The plot below depicts how the presence of in-app purchases differ between the top 10% and bottom 90% of the market. The majority (over 50%) of games in both groups contain in-app purchases. This differs from the trend we saw with the upfront cost of apps where most were free. Developers and the companies that produce the games are able to increase downloads by making the upfront cost of the app low. However, once consumers have downloaded the app and are invested in playing it, they are more likely to make purchases to enhance the gaming experience.



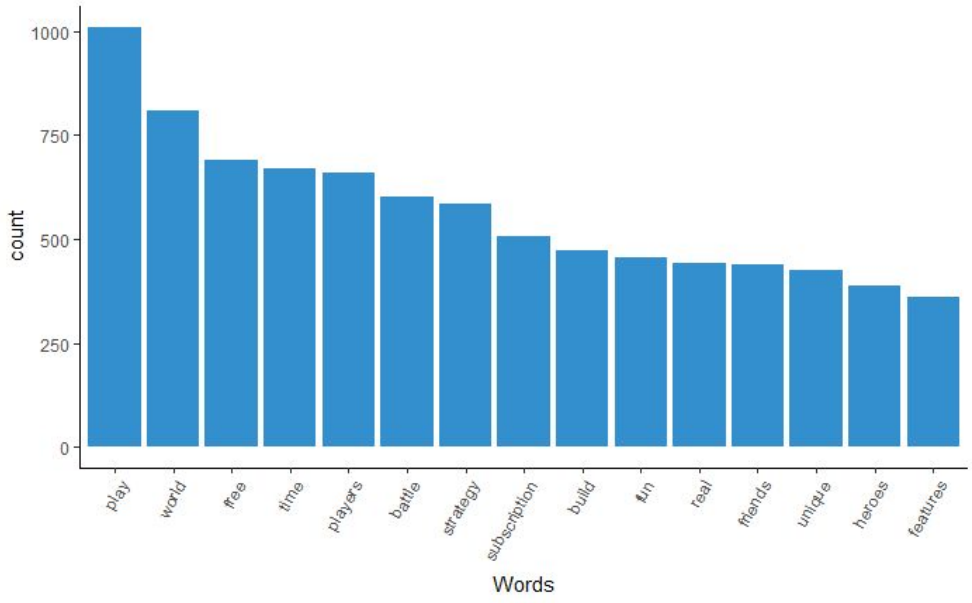
The same relationship can be seen below when the top 10% of strategy games were compared to the total market. We see, this time, that the proportion of the total market that has in-app purchases is over 60%. That is slightly higher than the bottom 90% of the market. This is likely due to the addition of the remaining 10% to the group.



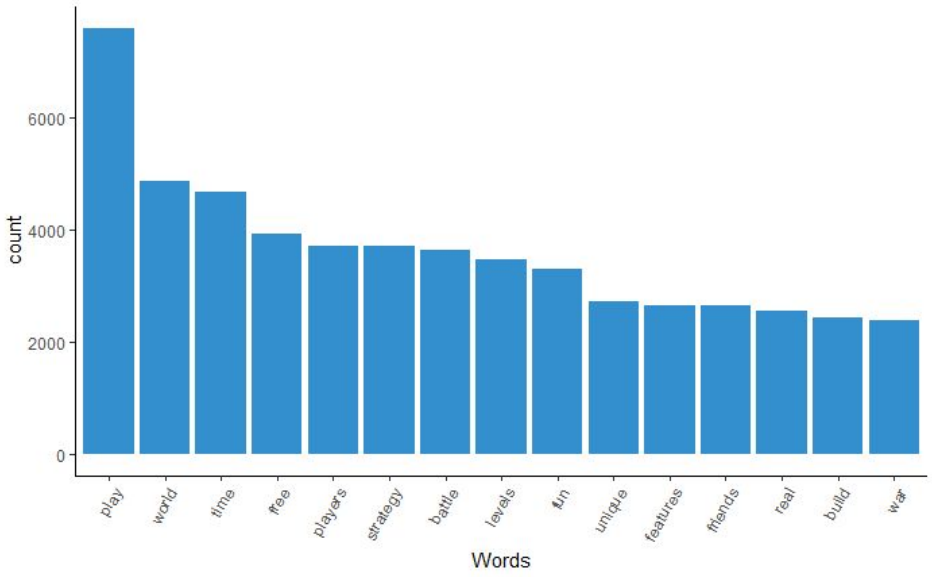
- **Description**

Our data set includes descriptions of the games that are included. We analyzed these descriptions using NLP. We separated words from the description variable and tried to find differences between All Games group and Top 10% Games group. Unfortunately, NLP didn't work well for this data set.

Here is the graph of the most frequently used words for top 10% of games:

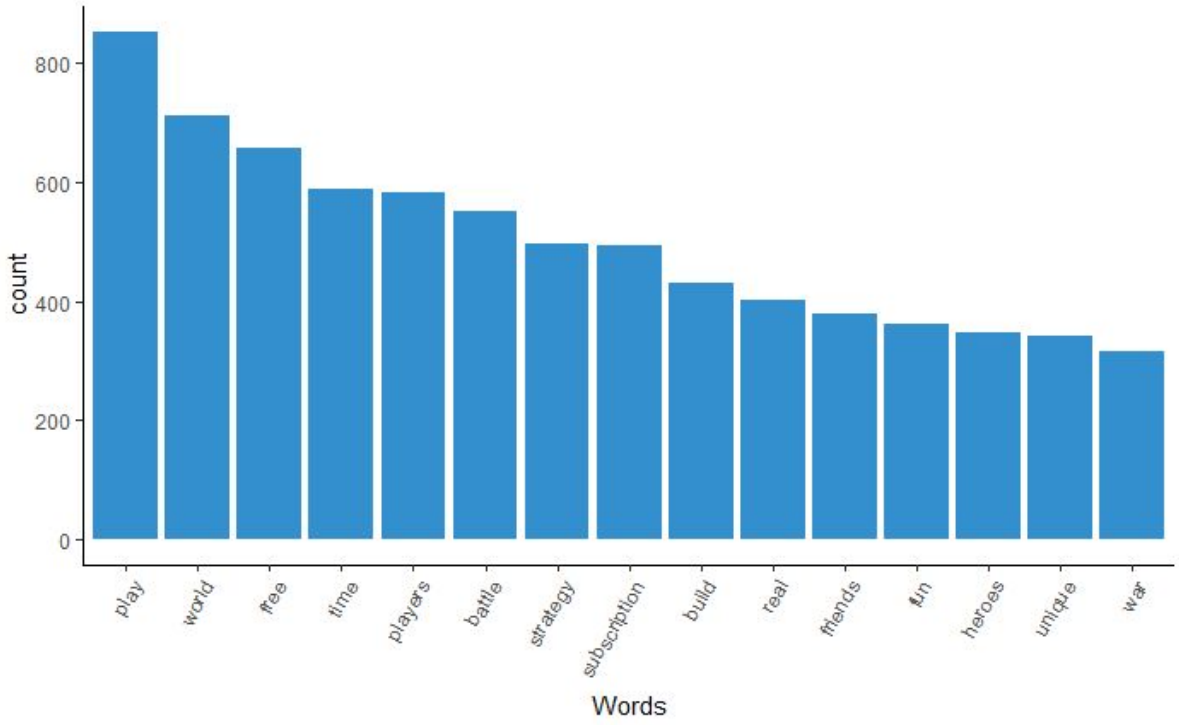


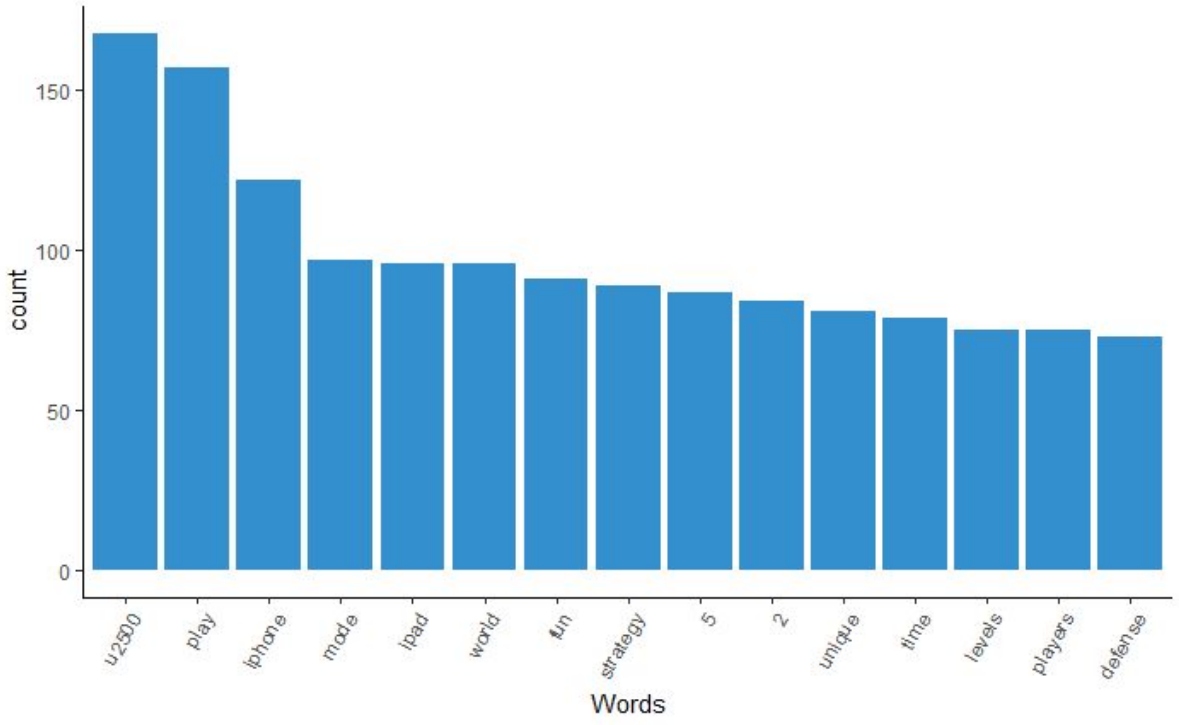
And here is the graph of the most frequently used words for all games:



There is no significant difference between the most frequently used words between the top 10% of the games and all games. We then made two graphs comparing word counts between free games and paid games in the top 10%. We wanted to check whether free games use a different strategy in description compared with paid games.

Here are the two graphs:





We were still unable to extract any useful information from the text analysis. NLP does not seem to be an effective way to explore our data set.

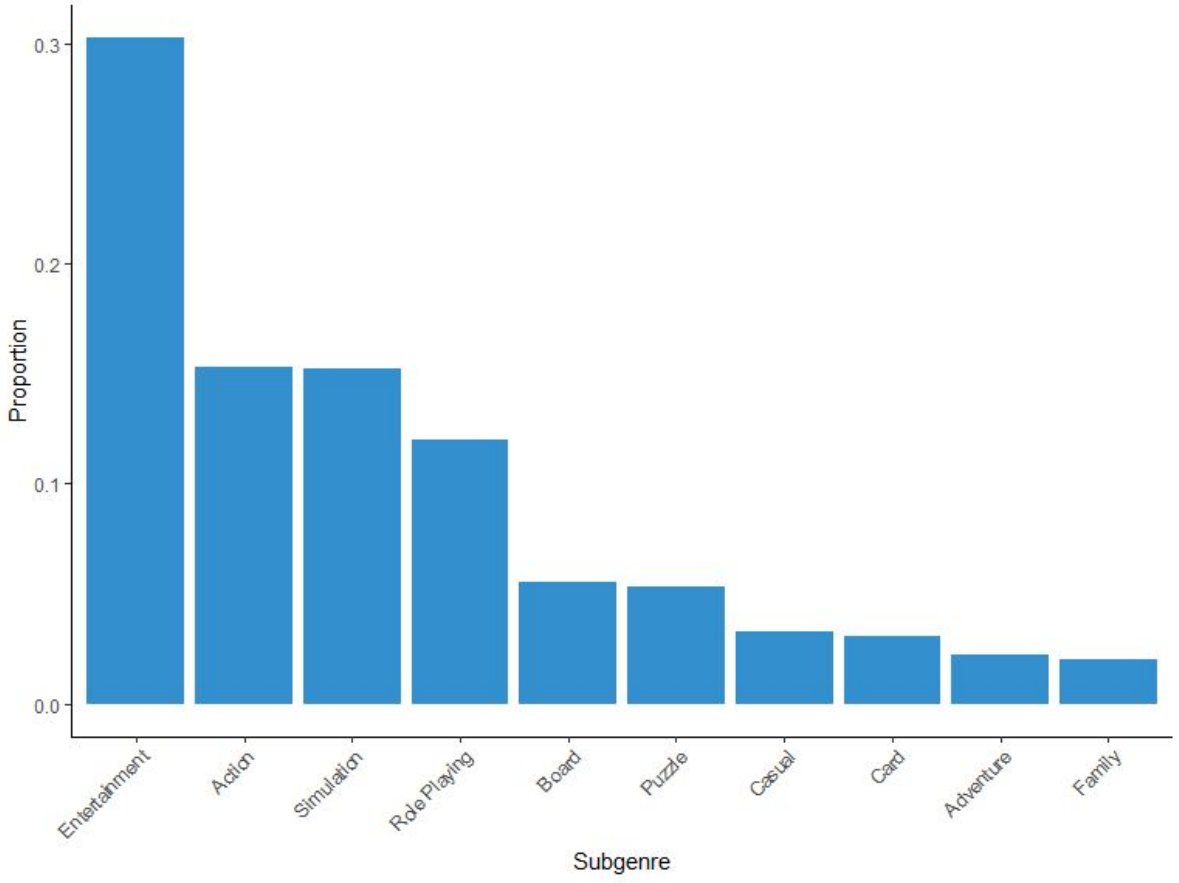
III. FURTHER EXPLORATION

Genre: Which genre is likely to succeed

The motivation for us to explore genre includes:

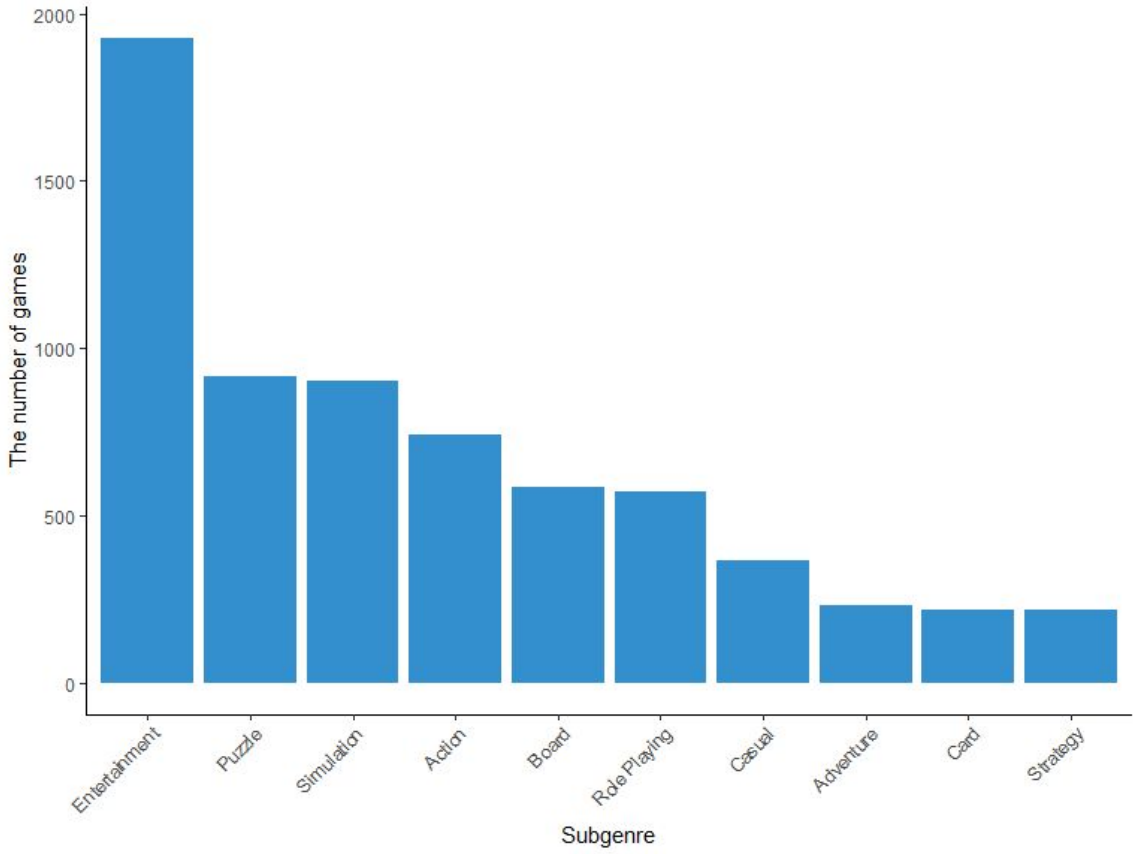
1) Most strategy games have main subgenres
2) There is significant difference among different subgenres
3) Competition within different subgenres could be different

First, we observed the top 10 subgenres within the top 10% of games:



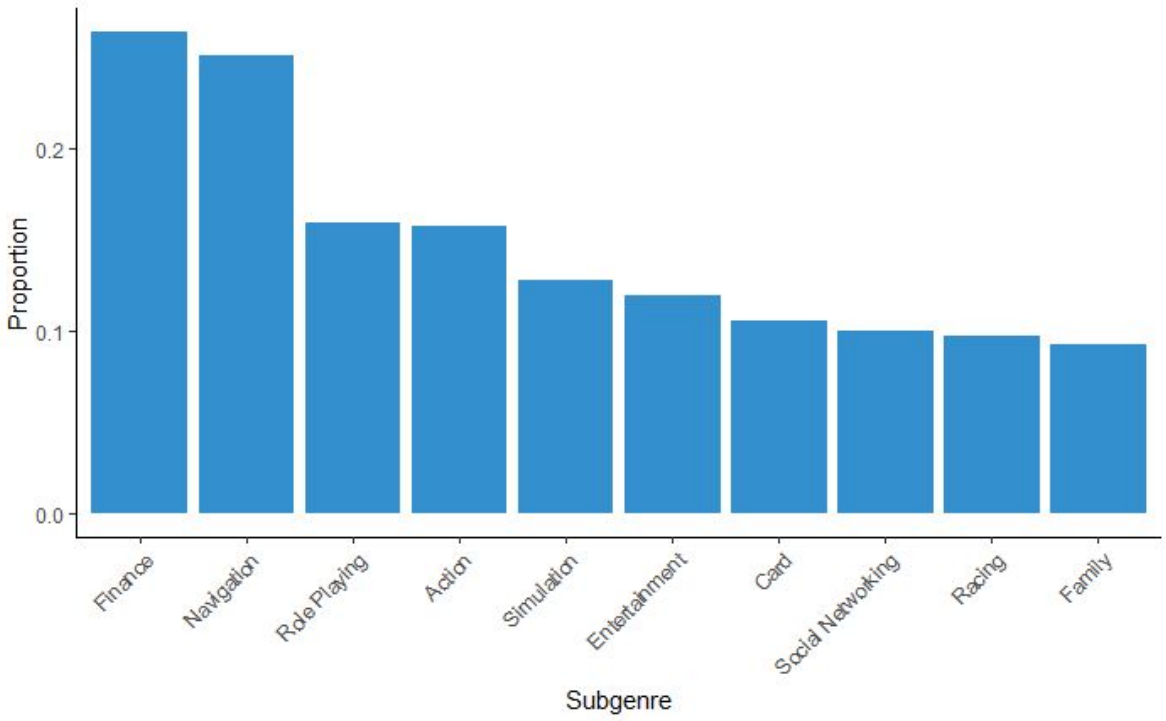
The plot shows that Entertainment is the most popular subgenre and has the largest consumer market. Action, Simulation, and Role Playing all have more than 10% in the top 10% group.

We wanted to see if any subgenre face really intense competition, so we checked the top 10 most intensely competitive subgenres among all the games.



As we can see from the plot, there are almost 2000 Entertainment games which have the most intense competition. Games with subgenres including Puzzle, Action, Simulation, Board, Role Playing, Casual, Adventure and Card also compete fiercely.

Family strategy games, which is the 10th popular subgenre, does not face as fierce competition as the previous subgenres. Therefore, we think checking the probability for a certain type of subgenre to be successful is a more rational way to get the ideal subgenres.



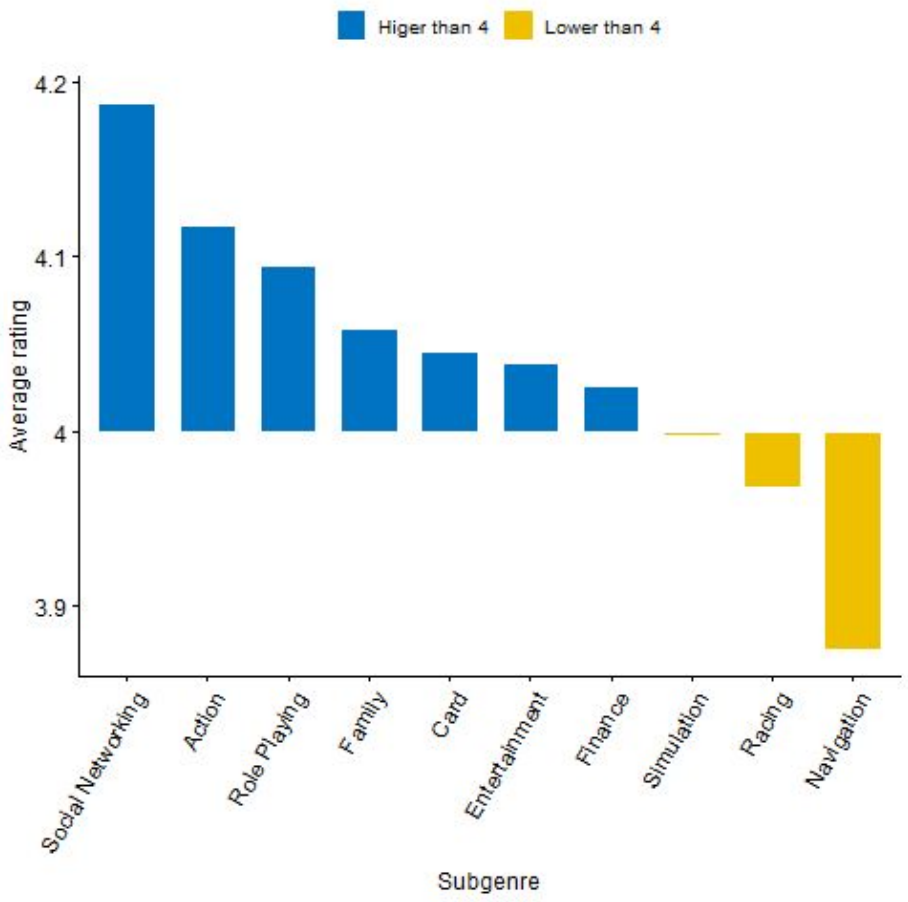
From all the three plots, we believe we can divide the subgenre choice into three strategies:

1) For large game developers or companies, developing Entertainment, Role Playing, Action, Simulation, Card games may be a good idea. Although these subgenres face intense competition, the probability of standing is still high, because large companies have enough funding for advertising and marketing. They may prefer to develop those subgenres because they have larger market share.
2) For sole developers, developing the above subgenres may not be a good idea because the competition is too intense and they may not have much money to spend on promotion. Therefore, Finance, Navigation, Racing, Social Networking would be good choices for them. Although these subgenres don't have large customer markets, they still have a high probability to stand out.
3) Developing Family games is a special strategy. It could be a good choice, but developing this kind of games may not lead to a huge success as it's not the most popular subgenre or has a high possibility to stand out.

Average user rating: A popular game with good reputation

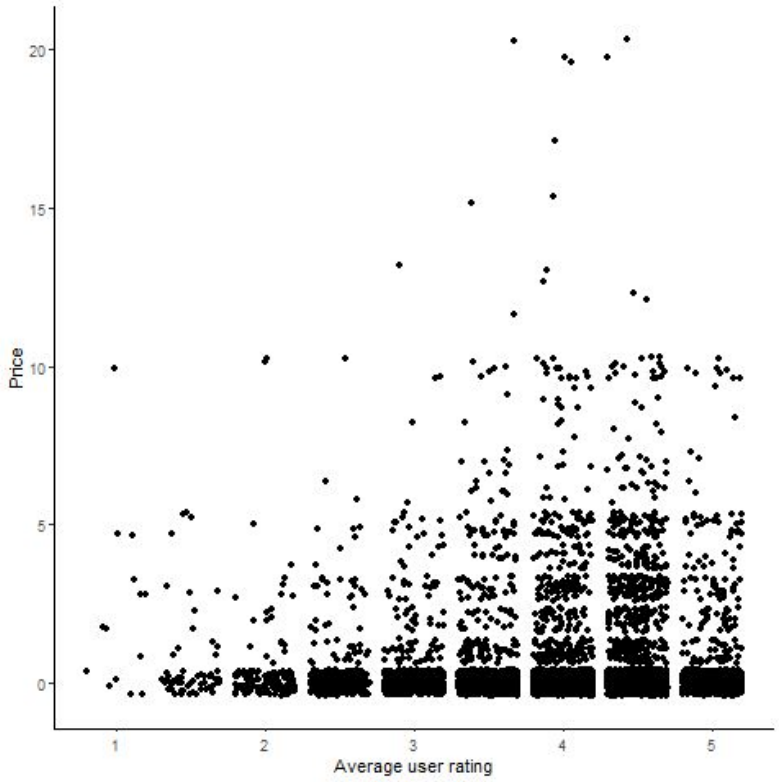
The motivation for us to explore quality is for developers who would like to go a step further and develop a game with high popularity as well as high ratings. Therefore, we use the average user rating as exploration object in this part.

First, we check the average user rating for subgenres that we concluded ideal for developers to develop.

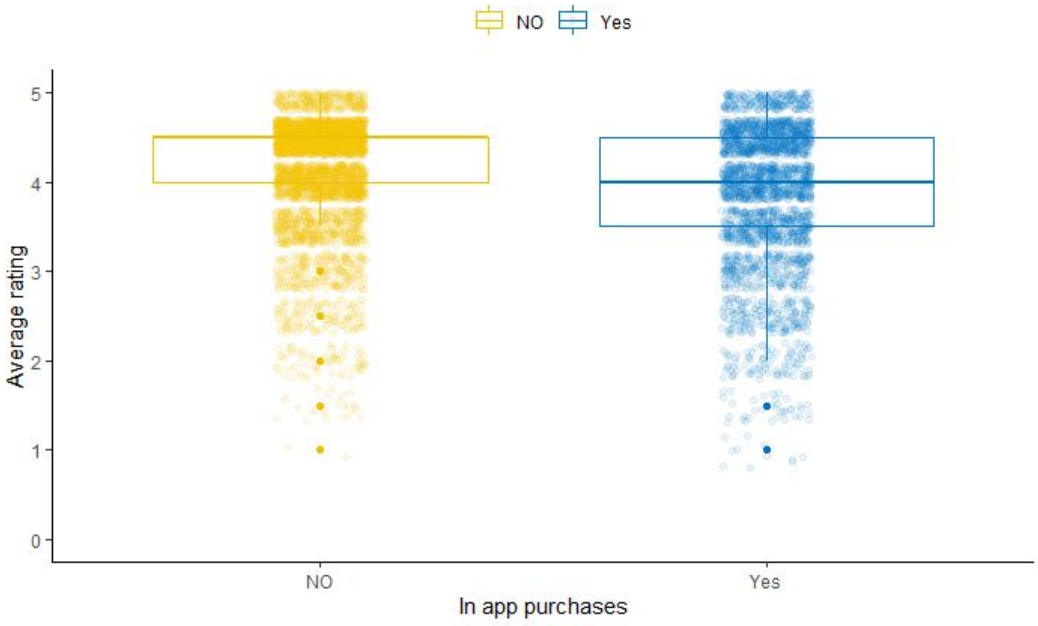


A rating of 4 out of 5 is the threshold for a high quality game. Therefore, Social Networking, Action, Role Playing, Family, Card, Entertainment and Finance are good choices as the average ratings are relatively high. However, Simulation, Racing and Navigation games did not perform as well.

We also checked the relationship between price and average ratings as well as the relationship between in app purchase and average ratings.



The first plot shows that customers tend to give paid games higher ratings while the free games are rated more unstable. This is potentially a result of consumers who are willing to pay more for an app being better informed about the quality of that app. It could also be the result of more expensive games not having as many ratings, so one person's rating would carry more weight.

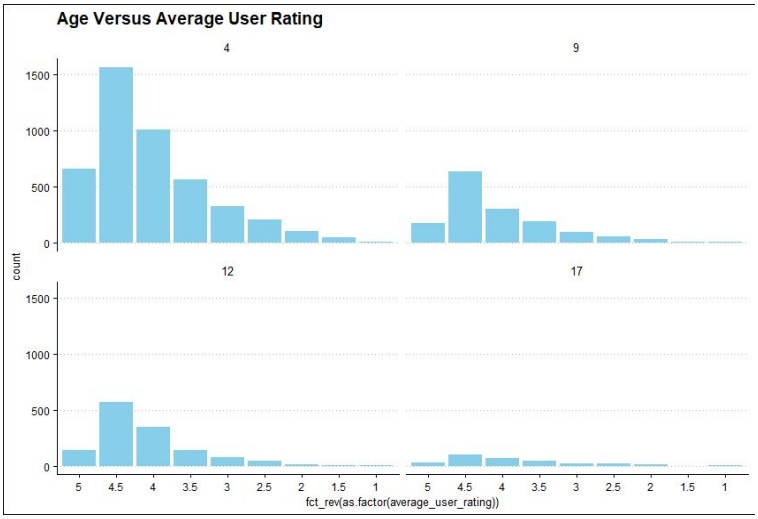


The second plot shows that in contrast with the price, games without in app purchases tend to be rated higher.

Therefore, developers may develop a paid game without in app purchase to get higher rating. This is interesting, because the plots from earlier suggested that creating a free game with in-app purchases was a more successful strategy.

Average User Rating: How it looks like within different age groups

The plot below explores the relationship between age group and the average user rating scores. The test is applied on the all-games group. Here, we divide the results into four groups according to age. For the four groups, most users rate a score of 4.5. There is a decreasing trend when the score is getting below 4.5 for each group. It seems like people are more likely to give a score of 5.0 when the game is rated as "4+", because, within the group of "4+", the number of people who give a score of 3.5 is less than the number of those who give a score of 5.0. Therefore, it may be better to come up with a game that falls within the "4+" group.



Basic modeling: GLM

Here we conduct a basic modeling for our data project. We would like to predict whether a strategy game is going to be successful or not.

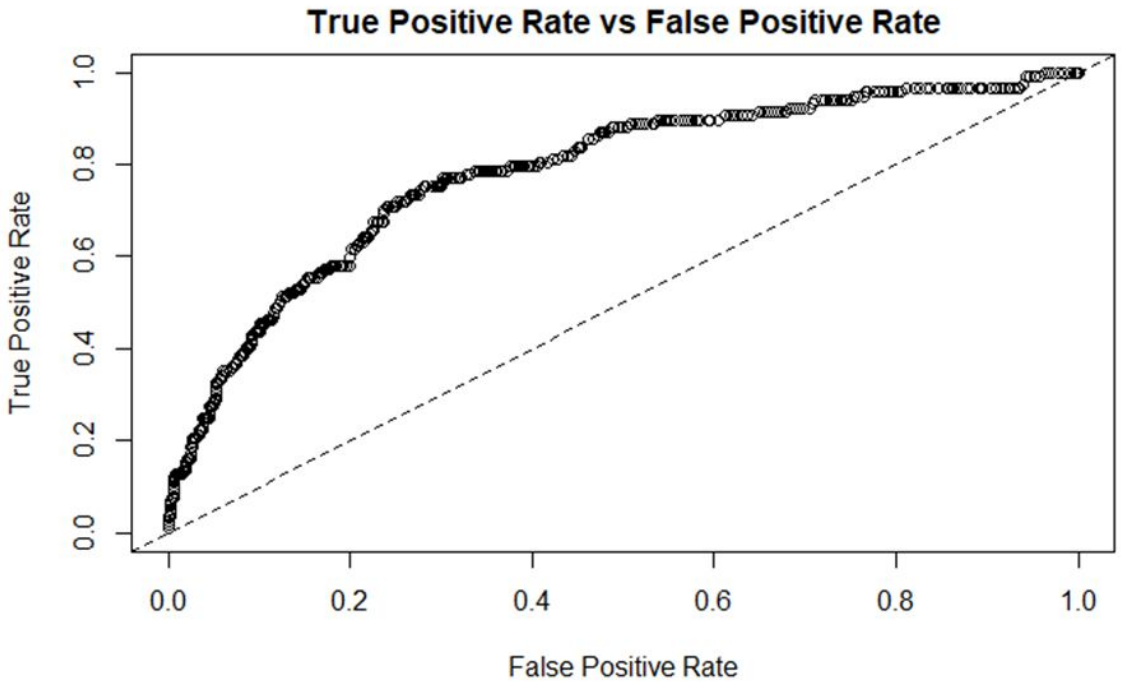
Response variable: Whether the game belongs to the group Top 10% number of ratings. Yi is of Bernoulli distribution.

Predictor variables:

1) Continuous variables: "price", "size_MB", "release_year"
2) Discrete variables: "age",
3) Dummy variables: "in_app_purchases", "ZH", "DE", "FR", "ES", "JA", "RU", "IT", "PT", "KO"

It is a binary classification problem, so we a use logistic regression model. The model we chose was a GLM model (family = "binomial"). We split the data into three parts: train, validation and test sets.

The result is shown with the plot below. We use AUC as the index for performance of this model for test data set.



IV. SUMMARY

Through our analysis, we concluded that a game in the Apple App Store will be positioned for success if factors like the size of the game and the language it supports taken into account. We discovered that the 250-450 MB game range is ideal for a game that features quality content and enhances the gaming experience. A popular game should also support at least 5 languages, as this is the trend for the Top 10% games on the app store.

Additionally, picking the right genre to is of the utmost importance since certain genres of games are generally more popular than others. We discovered that entertainment games are the most popular games within the the top 10% of games. For smaller publishers (i.e. start-ups), we recommend the family genre since it is popular yet less saturated than the entertainment genre.

Ethically, there is a need to ensure that the guidelines for establishing age ratings are followed, when publishing games. We discovered that there is an increasing trend in popularity for games with age ratings 9+ and 12+. This could mean that video game enthusiasts generally do not enjoy games that have a lot of violent content or games that have excessive sexual content.

Biography

G., Deyan. “61+ Revealing Smartphone Statistics For 2019.” TechJury, 28 March 2019, https://techjury.net/stats-about/smartphone-usage/#gref

Apple Inc. “App Review - App Store.” *Apple Developer*, https://developer.apple.com/app-store/review/.

Siegler, MG. “Here's How IPhone App Store Ratings Work. Hint: They Don't.” *TechCrunch*, TechCrunch, 29 June 2009, https://techcrunch.com/2009/06/29/heres-how-iphone-app-store-ratings-work-hint-they-dont/.

Liao, Sean. “Structure Your App.” *Migrating to Android for IOS Developers*, 2014, pp. 61–171., doi:10.1007/978-1-4842-0010-0_3.